economies

MDPI

*Article*

# Explaining Global Happiness: Evidence from Decision Trees and Necessary Condition Analysis

**Teresa Torres-Coronas [1] and Jorge de Andrés-Sánchez [2,*]**

[1] Social and Business Research Laboratory, Department of Business Management, Universitat Rovira i Virgili, Campus Sescelades, 43007 Tarragona, Spain; teresa.torres@urv.cat
[2] Social and Business Research Laboratory, Department of Business Management, Universitat Rovira i Virgili, Campus Bellissens, 43204 Reus, Spain
* Correspondence: jorge.deandres@urv.cat

**Abstract**

In this paper, the factors linked to happiness are analyzed on the basis of the World Happiness Report (WHR) model, introducing a methodological approach that differs from traditional econometric techniques. More specifically, this study examines how the core variables of the WHR model interact in relation to happiness and whether some of them also emerge as necessary conditions within the framework of necessary condition analysis (NCA) for attaining higher happiness levels. Using Gallup World Poll data for the 2022–2024 period, the Cantril Ladder is employed as a measure of subjective well-being, and gross domestic product (GDP) per capita, social support, healthy life expectancy, freedom, generosity, and perceived corruption are considered explanatory variables. This study makes two contributions. First, it applies a decision tree regression model to identify interactions among the correlates of happiness while also facilitating the classification of countries into homogeneous groups according to their well-being configurations. This approach improves interpretability relative to linear models because it does not require prior specification of those interactions. Second, this paper incorporates necessary condition analysis to distinguish between factors that are merely influential and those that emerge as necessary conditions for attaining certain levels of happiness. These assessments make it possible to identify minimum thresholds in key variables, introducing a necessary-condition logic. The results show that social support and GDP per capita emerge as the main structuring variables in the tree and are strongly associated with differences in happiness, whereas freedom emerges as a prominent condition in the NCA results. The findings also show that some factors with low correlation may still play a relevant role in specific contexts because of nonlinear effects and interactions. Overall, the results of this study offer an analytical reinterpretation of the WHR model by combining structural segmentation and threshold identification, advancing the understanding of happiness as a multidimensional, nonlinear phenomenon associated with specific configurations of factors.

**Keywords:** happiness; subjective well-being; Cantril ladder; cross-country analysis; machine learning; configurational approach; necessary condition analysis (NCA); Bottleneck analysis

Check for updates

Academic Editor: Periklis Gogas

Received: 31 March 2026
Revised: 29 April 2026
Accepted: 6 May 2026
Published: 11 May 2026

**Copyright:** © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

The pursuit of happiness has been, since the earliest stages of human thought, one of the fundamental aims of existence. In classical philosophy, Aristotle placed eudaimonia as

the highest good toward which all actions tend in his work Nicomachean Ethics (Turner, 2018). Centuries later, schools of thought such as utilitarianism, represented by Jeremy Bentham and John Stuart Mill, redefined happiness in terms of maximizing aggregate well-being, turning it into a normative criterion for social organization (Clarke, 2022). Even in the modern political sphere, happiness has been elevated to an explicit objective, as reflected in the United States Declaration of Independence, which recognizes the "right to the pursuit of happiness" as an inalienable principle (Leavitt, 2025). This historical persistence shows that happiness is not only an individual aspiration but also a collective ideal that has guided ethical, political, and social thought over time, as exemplified by the United Nations Sustainable Development Goals (Grinde, 2022).

This concern with happiness has led to systematic development in the field of economics, giving rise to what is known as the economics of happiness. This approach emerges as a response to the limitations of traditional indicators of well-being, such as gross domestic product, in adequately capturing people's quality of life (Agrawal et al., 2023). In this context, the Easterlin paradox highlights that increases in income do not necessarily translate into higher levels of happiness in the long term (Easterlin, 1974). Since then, the literature has incorporated subjective measures of well-being, relying on international surveys and cross-country comparable datasets, which has enabled a more comprehensive analysis of the factors associated with happiness and has generated a flourishing body of research (Veenhoven, 2025). Contributions from welfare economics and the capability approach have consistently reinforced the idea that human well-being should be understood in multidimensional terms, integrating factors such as health, social relationships, and individual freedom (Agrawal et al., 2023; Stiglitz et al., 2009).

Although happiness and well-being are often used interchangeably in the literature, they are far from being exact synonyms (Morse, 2026). In this study, we regard well-being, more specifically subjective well-being, as the broader umbrella concept, encompassing different evaluations that people make of their lives, whereas happiness is treated as a closely related but narrower evaluative expression within that broader domain (Vittersø, 2025). Accordingly, happiness is viewed here as an important indicator of well-being rather than as a concept fully equivalent to it. Consistent with the World Happiness Report (WHR) approach, the empirical analysis operationalizes happiness through the Cantril Ladder (CL) while acknowledging that this instrument captures a self-reported evaluation of one's life rather than a purely affective notion of happiness (Helliwell et al., 2025).

The economics of happiness has been found to be among the most influential empirical applications in the measurement framework developed from the Gallup World Poll and used by the WHR, which is published annually (Helliwell et al., 2025). Within this framework, the CL serves as the central measure of happiness. It is a scale from 0 to 10 on which individuals evaluate their current life by placing it between the worst and the best possible life for them. On the basis of this evaluative measure of subjective well-being, the literature identifies a set of factors associated with the observed differences across countries. The baseline WHR model, widely used in the literature, identifies GDP per capita, social support, healthy life expectancy, freedom to make life choices, generosity, and perceptions of corruption as the main associated factors (Helliwell et al., 2025). This framework, shown in Figure 1, fits squarely within the economics of happiness, as it combines subjective well-being measurement with economic, social, and institutional variables associated with why some societies exhibit higher levels of happiness than others do (Turner, 2018).

WHR models and CL measures have generated an extensive body of literature. On the one hand, this research explores the incorporation of new explanatory variables beyond the core factors, such as inequality and education (Kamińska et al., 2025; Zong, 2024) or religiosity (Organisation for Economic Co-operation & Development, 2015). On the

other hand, it has also examined the explanation of happiness from alternative conceptual frameworks, e.g., (Pereira et al., 2023). Additionally, methods other than conventional statistical techniques have been tested, such as neural networks or ensemble trees, to gain new insights into how WHR variables interact in relation to happiness levels (Çelik et al., 2025a; Chaudhary et al., 2020).

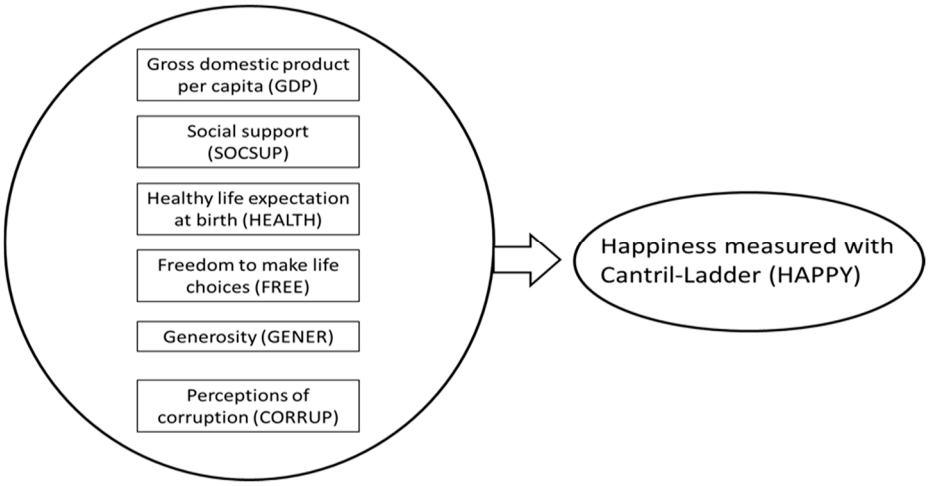



**Figure 1.** World Happiness Report baseline model. Source: Own elaboration from Helliwell et al. (2025).

The present study falls within this third line of research. Against this background, the central research question of this study is how the core variables of the WHR model interact in relation to happiness and whether some of them also emerge as necessary conditions for attaining higher happiness levels. More broadly, the study is guided by the expectation that happiness is associated with nonlinear configurations of WHR factors and with variable-specific threshold effects. Taking the baseline WHR model as a reference, two research objectives (ROs) are addressed through the sequential use of decision tree regression (DTR) and necessary condition analysis (NCA). Accordingly, we establish the following:

Research objective 1 (RO1) = To analyze how the core variables of the WHR model interact with different levels of happiness. Whereas in conventional statistical methods such as linear regression, interactions between variables must be specified a priori, as in Kamińska et al. (2025), DTR not only allows these interactions to be identified and visualized automatically from the available sample (Loh, 2011) but also enables the classification of countries into groups that exhibit not only a certain degree of homogeneity in happiness levels but also in the way explanatory variables interact in relation to happiness.

Research objective 2 (RO2) = To determine which WHR variables constitute necessary conditions for attaining higher levels of happiness. This is a relevant analysis, since when a factor is a necessary condition, it represents a potential bottleneck whose poor performance cannot be compensated for by improvements in other factors. This objective involves placing the evaluation of happiness within the logic of necessity, addressed through necessary condition analysis (NCA) (Dul, 2024a).

Notably, RO1 and RO2 are complementary objectives. While RO1 seeks to identify factors that operate as sufficient conditions—such that "if factor X is positively associated with happiness, higher values of X tend to be associated with higher happiness levels"—RO2 aims to determine whether such a factor is also necessary; that is, to establish the extent to which a statement such as "without a level of X, Xa, it is not possible to achieve a level of happiness Ha" holds true. Thus, the present study makes two contributions:

- First, it develops a DTR-based approach that allows for a nonlinear analysis of the interactions among the main WHR factors associated with happiness, identifying

specific combinations of factors associated with different levels of well-being across countries within the WHR conceptual framework. Although ensemble tree methods have already been applied to Gallup World Poll data (Çelik et al., 2025a), their primary purpose has been predictive. In contrast, this study, in addition to considering model fit, emphasizes understanding the interaction among explanatory variables and their joint association with national happiness, as well as their ability to group units of analysis into homogeneous clusters both in terms of happiness levels and in terms of how those levels are associated with different variable configurations. Furthermore, while DTR does not provide measures of statistical significance, it allows the establishment of a hierarchy of relative importance among factors by quantifying their contribution to the reduction in node impurity (Loh, 2011).

- Second, this study complements traditional regression-based analyses (e.g., Çelik et al., 2025b; Helliwell et al., 2025) by applying NCA, which makes it possible to identify which factors, beyond being associated with happiness levels, emerge as necessary conditions for attaining higher observed levels of happiness. In this way, the analysis is framed within a logic of necessity (Dul, 2024a). Despite the importance of determining whether a factor may act as a bottleneck for attaining higher happiness levels, to the best of our knowledge, this type of analysis has not yet been applied in the happiness literature.

This configurational perspective is theoretically appropriate for the study of happiness. Happiness is unlikely to be well captured by a single additive specification operating uniformly across countries; rather, it is more plausibly associated with combinations of economic, social, and institutional conditions whose relevance may vary across contexts. In multidimensional phenomena of this type, the effect of one factor may depend on the levels reached by others, and relevant thresholds may matter as much as average marginal associations. From this perspective, nonlinear and configurational approaches are not only empirically different from conventional linear models but also better aligned with the underlying structure of the phenomenon under study.

The use of decision tree regression does not imply that tree-based models are universally superior to conventional linear regression. Rather, it reflects a better match with the analytical objective of this study. Linear regression is appropriate when the main purpose is to estimate average net associations under an additive specification, but it requires nonlinearities and interaction terms to be specified ex ante. In contrast, decision tree regression is especially useful when the aim is to detect heterogeneous configurations, threshold effects, and hierarchical interactions directly from the data (Rad et al., 2025). In this sense, tree-based models provide a transparent and visually interpretable representation of how different combinations of predictors are associated with the outcome, revealing conditional patterns that may remain hidden in a standard linear specification (Loh, 2011). Prior methodological work has shown that regression trees can offer a predictive performance comparable to that of conventional regression in some settings while retaining the additional advantage of a tree structure that facilitates the interpretation of variable roles, thresholds, and interactions (Loh, 2011). This makes decision trees especially suitable for the present study, where the interest lies not only in average associations but also in identifying differentiated well-being regimes and the variable combinations associated with them.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature on the economics of happiness and the analysis of the factors associated with happiness. Section 3 describes the data and variables used. Section 4 presents the empirical methodology, including the decision tree model and necessary condition analysis. Section 5 reports the main results. Finally, Section 6 provides conclusions and outlines potential directions for future research.

## 2. Literature Review

### *2.1. Literature Expanding the Baseline World Happiness Report Model*

The first group of studies uses the baseline WHR model (Helliwell et al., 2025) while introducing additional variables. Thus, these studies expand the baseline WHR model with new elements or, alternatively, use the WHR explanatory variables as control variables that help assess the significance of the newly incorporated factors.

Various studies introduce technological, institutional, or cultural dimensions that enrich the understanding of well-being. When digitalization is incorporated, it maintains a positive relationship with happiness, although this relationship is heterogeneous across regions and has ambivalent effects depending on the institutional context (Ionescu-Feleagă et al., 2022). Likewise, the consideration of sustainability shows that the relationship between well-being and sustainable development is complex, nonlinear, and mediated by cultural and institutional factors (Detrinidad & López-Ruiz, 2024). In African countries, foreign direct investment is positively associated with happiness, although its contribution to well-being appears to depend on other structural and institutional factors within each country (Kariuki & Karanu, 2024).

Other studies introduce specific macroeconomic and political variables that affect well-being beyond income levels. A negative effect of the credit cycle on happiness has been found, although this effect is conditioned by factors such as social support or freedom (Li et al., 2020). Similarly, there are reports indicating that political and fiscal decentralization improve happiness, especially when spending is managed at the subnational level, although with nuances derived from citizens' tax perceptions (Lago et al., 2025).

Moreover, other studies focus on more specific institutional and social factors. Machani and Shtudiner (2026) experimentally analyze perceptions of corruption and conclude that they affect mainly negative emotions rather than overall happiness. When cultural gender norms are considered, precarious masculinity beliefs are negatively associated with national happiness through their impact on structural variables such as social support or health (Mahalik & Harris, 2026). Zong (2024) partially questions the traditional model by pointing to the limited relevance of generosity and proposing the inclusion of variables such as inequality or education.

Finally, some studies extend the analysis to applied economic domains. It has been observed that social happiness influences corporate decisions, such as corporate cash holdings, acting as a "soft institution" (Chen et al., 2025). Analysis of the role of human resource practices at the country level reveals that more flexible, inclusive, and personal development-oriented labor policies are associated with higher levels of aggregate well-being (Pološki Vokić & Klindžić, 2024).

### *2.2. Literature Using the Cantril Ladder Score as the Outcome and Explaining It Through Models Alternative to the World Happiness Report*

Another strand of the literature uses the CL measure of happiness but structures its explanation through analytical models that clearly diverge from those proposed by (Helliwell et al., 2025). In particular, some studies incorporate more cultural approaches, such as that by Hofstede (1980), while others redefine and measure variables such as health, wealth, or social support through alternative and conceptually distinct indicators.

One initial line of findings in this regard suggests that happiness tends to converge internationally: countries with lower initial levels of happiness improve more rapidly, and dispersion across countries declines over time (Ram, 2021). This points to a certain global equalization of well-being, although it does not by itself explain the specific mechanisms driving it.

The literature shows that there is no universal recipe for achieving high levels of well-being: income, inequality, institutional quality, and culture combine in different ways depending on the national context (Pereira et al., 2023). Thus, happiness responds to multiple and asymmetric configurations rather than simple linear relationships. Along the same lines, the role of culture has been highlighted: societies with lower power distance and greater individualism and indulgence generally exhibit greater happiness, whereas during the pandemic, long-term orientation favors better adaptation (Rajkumar, 2023). Similarly, it has been observed that women are generally as happy as or happier than men are and that this gap persists even after socioeconomic factors are controlled and does not depend on the level of development or institutional context (Zweig, 2015).

In these contributions, the economic dimension remains central, but it appears more nuanced. Greater income concentration among the richest 1% is associated with lower well-being, especially in Europe, where several emotional experiences also worsen (Powdthavee et al., 2017). In addition, both income and body mass index show a nonlinear relationship with happiness: moderate increases may improve well-being, but excessive increases ultimately reduce it. Moreover, inequality and unemployment have consistent negative effects on happiness levels (Kasman & Kasman, 2023).

Other studies broaden the focus toward less conventional structural factors. Latitude is positively related to quality of life and happiness, probably because of its association with historical trajectories, more effective institutions, and higher levels of human development (Pawliczek et al., 2022). Similarly, internal displacement caused by conflict deteriorates perceptions of peace but does not necessarily reduce national happiness, suggesting that happiness also depends on structural variables such as governance, employment, or food security (Collado, 2024).

### *2.3. Literature Investigating Analytical Methods Alternative to Conventional Ones*

The third line of studies analyzed does not necessarily introduce new variables or theoretical frameworks into the WHR scheme but instead proposes analytical alternatives that go beyond conventional statistical techniques, such as regression, which remains the dominant approach in most of the literature. These studies rethink the way relationships among already known variables are analyzed by exploring nonlinear, configurational, predictive, or hybrid structures. Thus, the advance lies not so much in what is explained but in how it is explained.

A clear first example is the proposal for a methodological respecification of the model through a multiplicative formulation, without incorporating new explanatory factors. This approach makes it possible to capture interactions among variables and diminishing returns, something that traditional linear models tend to represent only in a limited way (Tofallis, 2020).

Other studies also rely on tools linked to fuzzy-set theory. These approaches are useful for analyzing complex and context-dependent phenomena such as national happiness since they make it possible to study equifinality, causal asymmetry, degrees of membership, and interdependencies among factors (Pereira et al., 2023; Trinh & Nhieu, 2026).

Other contributions prioritize the visual and descriptive exploration of data over strict causal estimation. Visualization tools with Tableau, cluster analysis, descriptive statistics, and adapted CRISP-DM models have been used to represent spatial and temporal patterns of global happiness, summarize regional profiles, and identify associations among variables (Lee et al., 2019).

A promising approach is the use of machine learning tools to improve the predictive capacity of pure or derived WHR approaches. Thus, methods such as LASSO regression, random forest, and XGBoost have been used, combining penalized linear techniques with

more flexible algorithms (Akanbi et al., 2024). Other studies combine prediction and interpretation more ambitiously through the simultaneous use of econometric models, machine learning algorithms, and probabilistic causal structures. In this context, techniques such as ridge regression, decision trees, random forest, and XGBoost have been integrated (Çelik et al., 2025a; Timmapuram et al., 2023), as have deep neural networks and graph neural networks (Khder et al., 2022) or Bayesian networks aimed at modeling causal relationships among variables and analyzing conditional effects (Chaudhary et al., 2020). Moreover, hybrid strategies combining unsupervised and supervised learning have been developed, such as the use of K-means, random forest, and XGBoost to group countries and predict happiness levels, although these strategies do not assign a central role to the formal measurement of variable importance (Yang & Xie, 2025).

While some contributions focus on comparing predictive models and complement their evaluation with correlations, trees, and internal importance indicators (Timmapuram et al., 2023), others strengthen this line of research through combinations of clustering, classification, and interpretive techniques such as Gini importance, permutation importance, decision rules, ROC curves, or confusion matrices, with the aim of ranking variables and better understanding the differentiated profiles of countries according to their happiness levels (Çelik et al., 2025a; Khder et al., 2022).

In this study, the baseline WHR model is reinterpreted through analytical techniques alternative to traditional regressions. To model the CL measure, we use decision trees, which, although not entirely novel given that they have already been applied in some recent studies (Çelik et al., 2025a; Chaudhary et al., 2020; Timmapuram et al., 2023), are taken here one step further in terms of interpretation. This technique can be used as an alternative to the pooled linear regression applied to panel data (Çelik et al., 2025b; Helliwell et al., 2025) and can simultaneously group countries into homogeneous sets according to their happiness levels and the way these levels are associated with the explanatory variables of the WHR conceptual framework. What is entirely novel, however, is the use of NCA within the WHR framework to establish which explanatory variables act as necessary conditions, to what extent they do so, and how they are configured as such. Next, we introduce the database used and the method of analysis.

## 3. Materials and Methods

### *3.1. Description of Variables and Formulation of Hypotheses*

The data and variables used in this study are those from the Gallup World Poll employed in the WHR (Helliwell et al., 2025) for the 2022–2024 period. Their descriptive statistics are reported in Table 1.

**Table 1.** Descriptive statistics of the variables used in the analysis.

| Variable | CL | GDP | SOCSUP | HEALTH | FREE | GENER | CORRUP |
|---|---|---|---|---|---|---|---|
| Mean | 5.59 | 9.67 | 0.8 | 61.12 | 0.79 | 0.02 | 0.72 |
| Std. Dev. | 1.16 | 1.13 | 0.13 | 6.79 | 0.12 | 0.16 | 0.17 |
| Min | 1.28 | 7.23 | 0.23 | 43.25 | 0.23 | −0.32 | 0.15 |
| Max | 7.78 | 11.81 | 0.99 | 73.5 | 0.98 | 0.59 | 0.95 |
| N | 423 | 407 | 423 | 411 | 417 | 407 | 402 |

The dependent variable is happiness (HAPPY), operationalized through the CL. Specifically, individuals answer the following question: "Imagine a ladder with steps numbered from 0 to 10, where 0 represents the worst possible life and 10 the best possible life for you. On which step would you say you personally feel you stand at this time?" The variable used is the national average of these responses.

GDP per capita (GDP) is measured in terms of purchasing power parity and constant international dollars on the basis of data from the World Development Indicators. In some cases, estimates and projections are used to complete the most recent time series.

Social support (SOCSUP) is constructed from the proportion of individuals who respond affirmatively to the question of whether they have relatives or friends they can count on in times of need. This variable reflects the availability of social support networks at the individual level aggregated to the country level.

Healthy life expectancy at birth (HEALTH) is based on data from the World Health Organization, which combine information on life expectancy and health status. When complete data are not available, interpolation and extrapolation techniques are used to ensure temporal consistency.

Freedom to make life choices (FREE) is measured from the proportion of individuals who report being satisfied with their freedom to choose what to do with their lives, thus capturing individual autonomy.

Generosity (GENER) is operationalized as the residual from the regression of the national average response to the question "Have you donated money to a charity in the past month?" on GDP per capita.

Finally, perceptions of corruption (CORRUP) are measured as the proportion of individuals who perceive widespread corruption in government or business, reflecting the degree of institutional trust.

Thus, as confirmed in Helliwell et al. (2025), the following hypothesis is proposed:

**H1.** *GDP is positively related to HAPPY.*

**H2.** *SOCSUP is positively related to HAPPY.*

**H3.** *HEALTH is positively related to HAPPY.*

**H4.** *FREE is positively related to HAPPY.*

**H5.** *GENER is positively related to HAPPY.*

**H6.** *CORRUP is negatively related to HAPPY.*

### *3.2. Data Analysis*

The empirical analysis is structured into two complementary blocks. First, correlations among variables are analyzed, and DTR is fitted to identify nonlinear relationships between happiness and its determinants and to classify countries into homogeneous groups. The correlation analysis and the way in which the explanatory variables participate in the tree splits make it possible to assess, from two different perspectives, the extent to which hypotheses H1–H6 are supported. The aim of these analyses is to implement RO1. The development of RO2, which seeks to analyze the status of the WHR variables as necessary conditions, is carried out using NCA. In all the cases, the analysis is conducted using R 4.4.2 and the relevant packages/libraries. All analyses were run with fixed random seeds to ensure the reproducibility of the resampling-based procedures. Specifically, a seed was set before the 5-fold cross-validation, bootstrap stability analysis, and random forest estimation. The steps implemented are described below.

#### 3.2.1. Variables and Correlation Analysis

The dependent variable is HAPPY, measured using CL. The explanatory variables are GDP, SOCSUP, HEALTH, FREE, GENER, and CORRUP, which capture the dimensions of

income, social support, health, freedom, generosity, and perceived corruption, respectively. In all analyses, observations with missing values for the variables involved are removed beforehand. Listwise deletion was applied separately within each analytical block so that each model was estimated using the complete cases available for the variables involved in that specific procedure.

A first correlation analysis among variables is then conducted to visualize the sign of the relationship between HAPPY and the explanatory variables, which is expected to be positive for SOCSUP, HEALTH, FREE, and GENER and negative for CORRUP (Helliwell et al., 2025). It will also make it possible to visualize whether there are high correlations among the explanatory variables that should be taken into account in the interpretation of the decision tree estimated later.

#### 3.2.2. Fitting and Analysis of the Decision Tree

The main interest of DTR is explanatory, so complexity restrictions are incorporated. In particular, a minimum size of approximately 30 observations per terminal node is needed, and a maximum of 12 terminal nodes is imposed to obtain an interpretable structure. Tree selection is performed through a 5-fold cross-validation procedure implemented with caret, in which each fold served once as the validation subset while the model was trained on the remaining folds. For this purpose, an exhaustive hyperparameter grid is defined, consisting of the complexity parameter (cp) $\in$ {0.0025, 0.0050, 0.0075, 0.0100, 0.0125, 0.0150, 0.0200}, maximum tree depth (maxdepth) $\in$ {3, 4, 5, 6}, the minimum number of observations required to perform a split (minsplit) $\in$ {60, 80, 100}, and the minimum number of observations in terminal nodes (minbucket) $\in$ {25, 30, 35}. This resulted in the evaluation of 252 candidate tree specifications.

For each combination, a tree is estimated, and its cross-validation metrics are recorded as averages across the five validation folds: root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$). In addition, for each specification, the number of terminal nodes and the total number of nodes in the resulting tree are counted. The models are then ranked according to their cross-validated predictive performance, prioritizing the highest average cross-validated $R^2$ and, secondarily, lower average cross-validated RMSE and MAE values. Among them, only the trees that satisfy the interpretability restriction (terminal_nodes $\leq$ 12) are selected. The final tree is defined as the best feasible model within that subset.

To assess the quality of fit of the tree considered optimal, its $R^2$ is calculated, which should be comparable to those obtained through linear regression in previous contributions (Çelik et al., 2025b; Helliwell et al., 2025). To determine whether hypotheses H1–H6 are consistent with the sample and to establish the relative importance of the explanatory variables, the split rules of the fitted tree are subsequently analyzed. These are based on thresholds of the explanatory variables that distribute the observations contained in that node into two groups: those with a higher outcome level and those with a lower one. Although the primary split rule of the node is of greatest interest, surrogate splits are also relevant for this assessment. The sign postulated by each hypothesis is considered consistent with the splits of a node (primary and surrogate) when variables with an increasing relationship with happiness act in the split as a "ceiling" for observations with lower happiness levels within the partition (as in H1–H5) or as a "floor" when the postulated relationship is decreasing (H6).

The relative importance of the variables is obtained by aggregating the error reduction attributable to each across the different nodes of the tree. Its calculation takes into account not only the reduction in error when variables act as the primary split but also when they do so as surrogate splits.

This analysis was implemented in R 4.4.2 using the caret package for cross-validation and model selection, rpart for tree estimation, rpart.plot for tree visualization, dplyr for data handling, and ranger for the random forest robustness check.

#### 3.2.3. Classification of Countries into Terminal Nodes

To interpret the segmentation generated by the tree, countries are characterized using the mean values of GDP, SOCSUP, HEALTH, FREE, GENER, and CORRUP observed over the 2022–2024 period. The fitted tree rules are then applied to these three-year averages, assigning each country to a terminal node. From this classification, the number of countries, the mean level of happiness, the observed happiness range, and the list of countries included are obtained for each node. This makes it possible to interpret the nodes as structural happiness regimes.

This procedure is implemented in R using the dplyr package for data aggregation and transformation and rpart for applying the rules of the final tree and assigning each country to the terminal nodes.

#### 3.2.4. Robustness Checks

To assess the robustness of the tree-based results, three complementary checks are performed.

First, we compare the best feasible trees in the hyperparameter grid to verify whether the main splitting structure remains stable across similarly well-performing specifications. This comparison makes it possible to determine whether the substantive interpretation of the model depends on a single hyperparameter configuration or, instead, is preserved across alternative trees with comparable predictive performance and interpretable complexity.

Second, we conduct a bootstrap stability analysis with 500 replications. In each replication, the tree is re-estimated using the selected hyperparameters, and information is recorded on the frequency of the main splits, out-of-bag predictive performance, and variable-importance profiles. This procedure allows us to examine whether the core structure of the tree remains stable under repeated perturbations of the sample.

Third, we estimate a random forest model using 5-fold cross-validation and permutation-based variable importance exclusively as an external robustness check of the variable-importance hierarchy identified by the decision tree. For this purpose, a reduced hyperparameter grid is explored over mtry $\in$ {2, 3, 4, 6}, splitrule $\in$ {variance, extratrees}, and min.node.size $\in$ {3, 5, 10}. Among the candidate specifications, the preferred model is selected by minimizing the cross-validated RMSE, while higher $R^2$ and lower MAE are used as secondary criteria in the event of ties. This ensemble model is not used as a competing interpretive framework since the decision tree remains the preferred specification for visualization and substantive interpretation.

These robustness checks were implemented in R using caret for cross-validation and model selection, rpart for decision-tree estimation, dplyr for data handling, and ranger for random forest estimation and permutation-based variable importance.

#### 3.2.5. Necessary Condition Analysis (NCA)

NCA is implemented using HAPPY as the explained variable; GDP, SOCSUP, HEALTH, FREE, and GENER as positive correlates; and CORRUP as a negative correlate. The estimation of the effect size of the necessary condition status is carried out using two procedures for constructing the free disposal hull: the ceiling envelopment-free disposal hull (CE-FDH) method and the ceiling regression-free disposal hull (CR-FDH) method. In the latter case, $\tau = 0.95$ is set in the required quantile regression. In addition, a permutation test with 1000 replications is applied for significance testing, adopting a 5% significance level. This analysis is conducted in R using the NCA package.

The bottleneck table is produced only for CE-FDH and is constructed using the actual values of the variables involved so that the thresholds are expressed in their original units. In the case of the positive correlates, the table reports the minimum values required to achieve a given level of HAPPY; in the case of CORRUP, it reports the maximum admissible levels for achieving that outcome.

Because NCA evaluates each condition separately as a ceiling relation with the outcome, the corresponding bottlenecks are estimated in a bivariate manner rather than through an additive multivariate specification as in regression-based approaches. When several conditions are analyzed for the same outcome, these bivariate bottleneck results can be interpreted jointly as a set of variable-specific threshold relations, without implying a fully multivariate causal model (Hauff et al., 2021). However, several WHR variables are moderately to strongly correlated. Accordingly, the NCA results should be interpreted as condition-specific bottleneck relations observed in the data rather than as evidence that the conditions are empirically independent or causally isolated. In this sense, NCA identifies whether a given variable sets a minimum threshold for attaining a given happiness level, whereas the correlation analysis and the tree-based results provide the broader context of interdependence among variables. These analyses were implemented in R using the NCA package.

Package versions and the annotated R scripts, including the fixed seeds used throughout the empirical analysis, including the decision tree models, robustness checks, random forest validation, and necessary condition analyses, are publicly available in the Supplementary Materials.

#### 3.2.6. Adequacy of Sample Size

The available sample size is more than sufficient for the analytical methods implemented. In the case of decision trees, it is common to rely on traditional regression criteria (Stupak, 2024). With 6 explanatory variables and approximately 402 complete observations, the sample far exceeds both the heuristic rule of 10 observations per predictor and the criteria set out in Green (1991). N is greater than or equal to 50 plus 8 times the number of predictors for testing the overall model, and N is greater than or equal to 104 plus the number of predictors for testing individual predictors, since in our case, these thresholds would imply 98 and 110 observations, respectively. In any case, additional complexity restrictions and minimum terminal node size constraints were imposed on the decision tree, ensuring a sufficiently parsimonious and interpretable model.

With respect to NCA, the sample size is sufficient to detect medium-sized effects or larger ($d > 0.1$), which are usually considered empirically relevant (Dul et al., 2023). In particular, for $d = 0.10$ and with 400 observations, the power of the permutation test reaches 100% in the case of CE-FDH and 0.81 in the case of CR-FDH under the assumption of a uniform distribution of HAPPY and the explanatory variable (Dul, 2024b).

## 4. Results

### *4.1. Results of the Correlation Analysis and Decision Tree Fitting (Research Objective 1)*

#### 4.1.1. Correlations, Tree Structure, and Variable Importance

Table 2 presents the correlation matrix for the variables analyzed. In particular, when the relationship between each variable and HAPPY is examined, the sign of the correlations is as expected: positive in all the cases except for CORRUP, which shows a negative correlation. Similarly, most of the correlations with HAPPY can be considered moderate or strong, with absolute values above 0.4. The exception is GENER, whose correlation with HAPPY is less than 0.1 in absolute value and can therefore be classified as negligible (Akoglu, 2018). With the exception of the correlation of GENER with the other variables,

all the correlations are significant for alpha < 0.01. Thus, the correlations support all the hypotheses except H5.

**Table 2.** Pearson's correlation matrix.

| | HAPPY | GDP | SOCSUP | HEALTH | FREE | GENER | CORRUP |
|---|---|---|---|---|---|---|---|
| HAPPY | 1 | | | | | | |
| GDP | 0.767 ** | 1 | | | | | |
| SOCSUP | 0.800 ** | 0.681 ** | 1 | | | | |
| HEALTH | 0.670 ** | 0.688 ** | 0.639 ** | 1 | | | |
| FREE | 0.577 ** | 0.321 ** | 0.523 ** | 0.474 ** | 1 | | |
| GENER | 0.068 | −0.082 | −0.015 | −0.073 | 0.063 | 1 | |
| CORRUP | −0.445 ** | −0.427 ** | −0.232 ** | −0.393 ** | −0.300 ** | −0.158 * | 1 |

Note: "**" denotes significance at $p < 0.001$, and "*" denotes significance at $p < 0.01$.

The best-performing tree among the 252 candidate specifications evaluated was obtained with the following parameters: cp = 0.0025, maxdepth = 4, minsplit = 60, and minbucket = 25. This model showed satisfactory predictive performance under 5-fold cross-validation, with average validation-fold values of RMSE = 0.644, $R^2$ = 0.702, and MAE = 0.4808. These metrics correspond to the average performance obtained across the validation folds of the cross-validation procedure; no additional independent hold-out validation set was used. In addition, the tree complexity analysis revealed, for the final selected value, a relative cross-validated error (xerror) of 0.3173 and a standard deviation (xstd) of 0.0273, suggesting a good balance between fit and parsimony. In the full sample, the tree also reaches an RMSE = 0.5721, an MAE = 0.4261, and an $R^2$ = 0.7613, which is consistent with its good in-sample fit and with a coefficient of determination similar to those reported in Çelik et al. (2025b) or Helliwell et al. (2025). Its structure is shown in Figure 2.

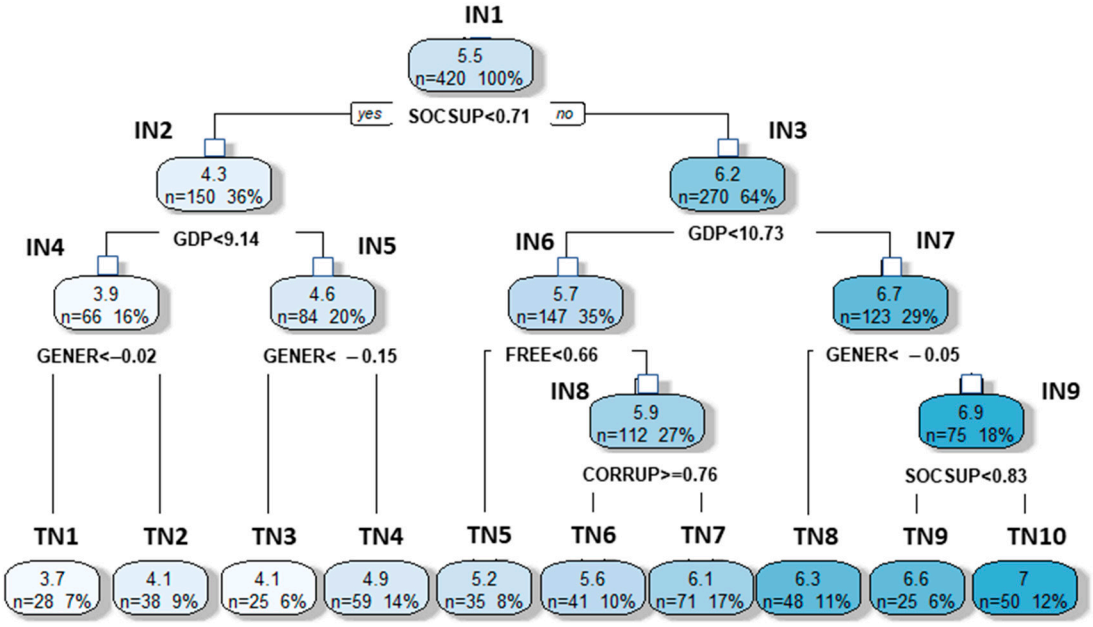



**Figure 2.** Decision tree fitted to HAPPY. Note: IN1, IN2, . . ., IN9 represent intermediate nodes, and TN1, TN2, . . ., TN10 represent terminal nodes.

With respect to the relative importance of the variables, the data in Figure 2 suggest that SOCSUP and GDP are the main structuring variables in the tree since they act as splitting criteria at the upper levels and organize the formation of the terminal groups of

observations. First, the tree is split on SOCSUP, indicating that this dimension functions as the most relevant initial threshold for differentiating happiness levels. GDP then enters as the second key variable in both resulting major branches, reinforcing its role as a central partitioning axis. Beyond these main splits, other variables, such as FREE, GENER, and CORRUP, refine the segmentation at deeper levels, playing a complementary role in refining the segmentation associated with happiness. Notably, GENER, despite having a virtually zero correlation with HAPPY, appears three times as a primary split. The absence of HEALTH from the primary splits should also be emphasized. Despite its low correlation with happiness, the splitting frequency of GENER reflects its role in nonlinear effects and conditional interactions, whereas HEALTH, although highly correlated with HAPPY, loses marginal relevance because of its collinearity with variables such as GDP and SOCSUP.

By examining the thresholds associated with each explanatory variable in the splits—both when they act as primary splits and when they are considered alternative splits—it is also possible to identify the sign of the relationship between HAPPY and the WHR variables. As shown in Figure 2 (where only the primary split is displayed) and in Table 3 (which also reports the alternative splits), in the case of GDP, SOCSUP, HEALTH, FREE, and GENER, the cutoff points that generate the partitions define values that should not be exceeded to fall into the group with lower levels of happiness. In contrast, in the case of CORRUP, the thresholds indicate values that must be exceeded to belong to the group of countries with lower levels of happiness within the partition. These findings are consistent with the hypothesized signs of the relationships between the explanatory variables and HAPPY proposed in Section 3.1.

**Table 3.** Primary and surrogate splits of the decision tree.

| NODE | Primary Split | Alternative Splits (Condition to Be in the Set with Lower HAPPY) |
|---|---|---|
| IN1 | SOCSUP < 0.71 | GDP < 10.31; HEALTH < 58.38; FREE < 0.69; CORRUP $\geq$ 0.48 |
| IN2 | GDP < 9.14 | SOCSUP < 0.48; FREE < 0.51; HEALTH < 52.20; GENER < $-$0.16 |
| IN3 | GDP < 10.73 | CORRUP $\geq$ 0.62; SOCSUP < 0.79; HEALTH < 58.93; FREE < 0.75 |
| IN4 | GENER < $-$0.02 | FREE < 0.60; SOCSUP < 0.53; HEALTH < 52.25; CORRUP $\geq$ 0.71 |
| IN5 | GENER < $-$0.15 | FREE < 0.61; SOCSUP < 0.64; CORRUP $\geq$ 0.66; HEALTH < 57.31 |
| IN6 | FREE < 0.66 | GDP < 10.19; SOCSUP < 0.85; HEALTH < 59.24; CORRUP $\geq$ 0.69 |
| IN7 | GENER < $-$0.05 | CORRUP $\geq$ 0.61; FREE < 0.66; SOCSUP < 0.83; GDP < 10.97 |
| IN8 | CORRUP $\geq$ 0.76 | GDP < 10.32; SOCSUP < 0.79; FREE < 0.79; HEALTH < 59.81 |
| IN9 | SOCSUP < 0.83 | CORRUP $\geq$ 0.29; GENER < 0.11; GDP < 11.06; FREE < 0.77 |

Note: The condition "<" or "$\geq$" refers to the observations classified in the nodes with lower levels of happiness.

The variable importance shown in Figure 3 confirms that SOCSUP is the variable with the greatest relative importance in the tree-based analysis, accounting for approximately 34.6% of the total importance, followed by GDP with 27.7%. Both act as the main structural axes of the tree, which is consistent with their role in the upper splits of the model. At the second level is HEALTH, with a contribution close to 19.8%, which highlights its substantive relevance in explaining well-being. Notably, although HEALTH does not appear as a splitting variable in the main nodes of the tree, it is highly important. This is because its contribution materializes indirectly through improvements in the predictive capacity of the model across multiple nodes, including possible surrogate splits, reflecting its cross-cutting role in the tree-based explanation of happiness. In contrast, FREE, GENER, and CORRUP have lower relative importance and play more specific or complementary roles in the final segmentation of the tree.

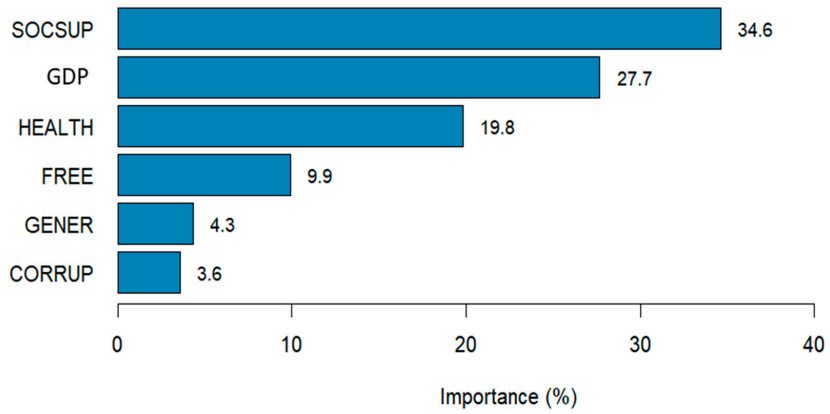



**Figure 3.** Variable importance in the explanation of happiness (values are proportions).

#### 4.1.2. Results of Robustness Checks

To assess the robustness of the tree-based results, three complementary checks were performed.

In the first robustness check, among the 252 candidate tree specifications defined in the hyperparameter grid and ranked according to their 5-fold cross-validation performance, we considered the ten best-performing feasible specifications satisfying the interpretability restriction of no more than 12 terminal nodes. This set included the specification selected for the main analysis in Section 4.1.1 and the next nine best-performing alternatives. These ten specifications showed very similar results in the cross-validation evaluation: $R^2$ ranged from 0.698 to 0.702, and RMSE ranged from 0.644 to 0.647. In all cases, the upper structure of the tree was identical: social support (SOCSUP) consistently appeared as the root split, whereas GDP per capita (GDP) defined both second-level partitions. These results indicate that the main substantive interpretation of the tree is stable across the best feasible specifications.

The second check consisted of a bootstrap stability analysis. This confirmed that the main structure of the decision tree is highly robust. As shown in Table 4, SOCSUP appears as the root split in 96.8% of the 500 bootstrap replications, whereas GDP dominates the second level of the tree in 72.7% of the cases. This finding indicates that the upper structure of the model remains largely unchanged under repeated perturbations of the sample. In addition, the data in Table 5 show that the out-of-bag predictive performance is stable across the replications (mean RMSE = 0.6573, mean MAE = 0.4964, mean $R^2$ = 0.6905), whereas the average number of terminal nodes remains very close to that of the selected tree (10.204, sd = 0.7639). Finally, the data in Table 6 indicate that the variable-importance hierarchy is highly consistent across bootstrap samples, with SOCSUP, GDP, and HEALTH systematically emerging as the three most important predictors. Overall, these results suggest that although some lower-level details may vary across replications, the substantive interpretation of the decision tree is robust.

In the third robustness check, it was observed that the preferred random forest specification was the model with mtry = 4, split rule = extratrees, and min.node.size = 3. This specification achieved strong predictive performance under 5-fold cross-validation (RMSE = 0.4880, $R^2$ = 0.8294, and MAE = 0.3705). The corresponding permutation-based variable importance is presented in Figure 4. The ranking is highly consistent with that obtained from the decision tree: SOCSUP and GDP remain the two most important predictors, and HEALTH continues to occupy the third position. Hence, the random forest results confirm that the main variable-importance hierarchy does not depend on a single-tree specification.

**Table 4.** Bootstrap stability of the decision tree (frequency of main splits).

| Split Position | Variable | Frequency (%) |
|---|---|---|
| Root split | SOCSUP | 96.8 |
| Root split | GDP | 3.2 |
| Second level | GDP | 72.7 |
| Second level | SOCSUP | 19.7 |
| Second level | FREE | 5.3 |
| Second level | CORRUP | 1.6 |
| Second level | GENER | 0.4 |
| Second level | HEALTH | 0.3 |

**Table 5.** Bootstrap stability of the decision tree (out-of-bag performance and complexity).

| Statistic | Mean | SD |
|---|---|---|
| RMSE (Out of Bag) | 0.6573 | 0.0425 |
| MAE (Out of Bag) | 0.4964 | 0.0309 |
| $R^2$ (Out of Bag) | 0.6905 | 0.0348 |
| Terminal nodes | 10.204 | 0.7639 |

**Table 6.** Bootstrap stability of the decision tree (relative variable importance across bootstrap samples).

| Variable | Mean | SD |
|---|---|---|
| SOCSUP | 0.3387 | 0.0248 |
| GDP | 0.2697 | 0.0199 |
| HEALTH | 0.1931 | 0.0136 |
| FREE | 0.115 | 0.0178 |
| CORRUP | 0.0459 | 0.0187 |
| GENER | 0.0376 | 0.0103 |

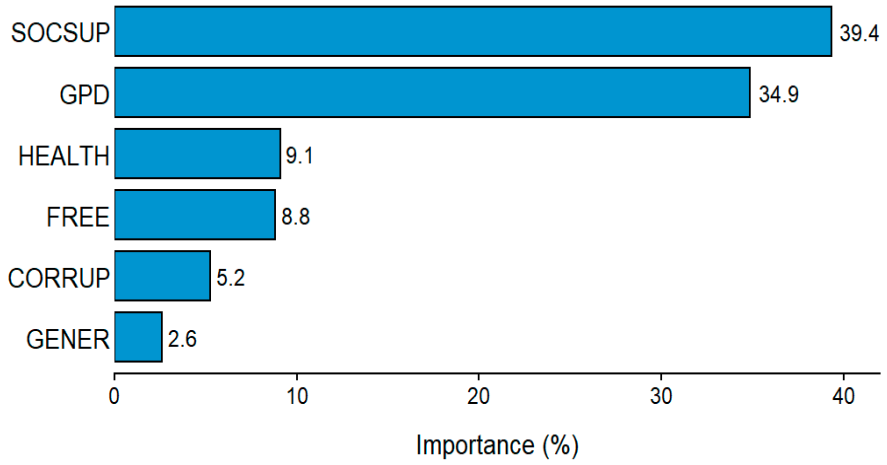



**Figure 4.** Variable importance in the explanation of happiness with random forest (values are proportions).

#### 4.1.3. Classification of Countries into Terminal Nodes and Structural Happiness Regimes

Table 7 shows the locations of the countries analyzed across the terminal nodes of the classification tree in Figure 2. This classification is based on the average values of the explanatory variables over the 2022–2024 period. At the lower end, terminal node (TN) 1 (TN1) mainly includes low-income Sub-Saharan African countries and Afghanistan. These are economies with very low social support, low income, and poor health indicators, which is consistent with this group concentrating the lowest levels of happiness. TN2 maintains a similar profile, dominated by Sub-Saharan Africa, although with slightly higher levels of social support and generosity, allowing for a modest improvement in average happiness.

**Table 7.** Country composition of the terminal nodes of the tree in Figure 2.

| Node | Descriptive Statistics | Countries |
|---|---|---|
| TN1 | HAPPY: 3.615 (1.567–4.540) GDP: 8.764 (8.437–9.099) SOCSUP: 0.439 (0.230–0.601) HEALTH: 50.957 (48.345–54.025) FREE: 0.554 (0.230–0.700) GENER: −0.074 (−0.134–−0.017) CORRUP: 0.653 (0.505–0.728) | Niger, Benin, Mali, Togo, Comoros, Malawi, Zimbabwe, Afghanistan |
| TN2 | HAPPY: 4.088 (3.059–5.071) GDP: 8.704 (8.347–9.077) SOCSUP: 0.568 (0.478–0.654) HEALTH: 50.806 (46.228–54.801) FREE: 0.619 (0.409–0.764) GENER: 0.095 (0.001–0.335) CORRUP: 0.690 (0.493–0.792) | Mozambique, Guinea, Burkina Faso, Chad, Gambia, Liberia, Madagascar, Ethiopia, Zambia, Tanzania, DR Congo, Sierra Leone |
| TN3 | HAPPY: 4.425 (2.691–6.324) GDP: 10.005 (9.720–10.275) SOCSUP: 0.628 (0.456–0.692) HEALTH: 56.678 (47.739–60.321) FREE: 0.627 (0.371–0.840) GENER: −0.240 (−0.320–−0.165) CORRUP: 0.675 (0.507–0.811) | El Salvador, Georgia, Gabon, Morocco, Tunisia, Jordan, Egypt, Botswana, Lebanon |
| TN4 | HAPPY: 4.770 (3.947–5.944) GDP: 9.469 (9.157–10.185) SOCSUP: 0.593 (0.374–0.688) HEALTH: 54.832 (49.781–62.544) FREE: 0.685 (0.504–0.873) GENER: 0.038 (−0.076–0.487) CORRUP: 0.714 (0.577–0.822) | Honduras, Albania, Nepal, Lao PDR, Congo, Côte d'Ivoire, Iraq, Nigeria, Cameroon, Senegal, Pakistan, Mauritania, Kenya, Ghana, Cambodia, Myanmar, India, Bangladesh |
| TN5 | HAPPY: 5.283 (4.739–5.838) GDP: 9.929 (7.230–10.700) SOCSUP: 0.787 (0.726–0.882) HEALTH: 57.630 (49.2–64.4) FREE: 0.563 (0.370–0.655) GENER: −0.071 (−0.299–0.159) CORRUP: 0.730 (0.585–0.831) | Greece, Russian Federation, Mongolia, Venezuela, Algeria, North Macedonia, South Africa, Iran, Türkiye, State of Palestine, Namibia |
| TN6 | HAPPY: 5.479 (4.018–6.306) GDP: 10.010 (8.839–10.707) SOCSUP: 0.790 (0.731–0.870) HEALTH: 59.052 (53.567–62.661) FREE: 0.737 (0.663–0.856) GENER: 0.013 (−0.199–0.502) CORRUP: 0.802 (0.767–0.842) | Panama, Thailand, Bosnia and Herzegovina, Republic of Moldova, Jamaica, Kyrgyzstan, Bolivia, Colombia, Peru, Bulgaria, Indonesia, Ukraine, Uganda, Sri Lanka |
| TN7 | HAPPY: 6.072 (5.294–6.914) GDP: 10.180 (9.193–10.618) SOCSUP: 0.793 (0.727–0.868) HEALTH: 59.754 (54.333–63.406) FREE: 0.773 (0.671–0.868) GENER: −0.063 (−0.222–0.205) CORRUP: 0.658 (0.409–0.762) | Costa Rica, Mexico, Uruguay, Kosovo, Serbia, Chile, Brazil, Nicaragua, Guatemala, Kazakhstan, Argentina, Uzbekistan, Viet Nam, Malaysia, Paraguay, China, Philippines, Mauritius, Montenegro, Ecuador, Dominican Republic, Armenia, Tajikistan |
| TN8 | HAPPY: 6.278 (5.938–6.770) GDP: 10.867 (10.739–11.020) SOCSUP: 0.831 (0.743–0.897) HEALTH: 63.373 (58.907–68.012) FREE: 0.693 (0.585–0.815) GENER: −0.147 (−0.287–−0.074) CORRUP: 0.691 (0.472–0.840) | Lithuania, France, Taiwan Province of China, Saudi Arabia, Romania, Poland, Spain, Italy, Slovakia, Latvia, Japan, Cyprus, Republic of Korea, Portugal, Hungary, Croatia |
| TN9 | HAPPY: 6.467 (5.325–7.127) GDP: 11.255 (10.980–11.650) SOCSUP: 0.792 (0.737–0.823) HEALTH: 64.623 (59.173–69.916) FREE: 0.779 (0.615–0.842) GENER: 0.062 (−0.043–0.219) CORRUP: 0.374 (0.081–0.638) | Luxembourg, Austria, United Kingdom, United Arab Emirates, Singapore, Bahrain, Hong Kong SAR of China |
| TN10 | HAPPY: 7.015 (6.283–7.736) GDP: 11.093 (10.838–11.580) SOCSUP: 0.875 (0.838–0.929) HEALTH: 64.806 (59.588–66.778) FREE: 0.800 (0.669–0.873) GENER: 0.073 (−0.048–0.190) CORRUP: 0.391 (0.116–0.756) | Finland, Denmark, Iceland, Sweden, Israel, Netherlands, Norway, Switzerland, Australia, New Zealand, Canada, Belgium, Ireland, Czechia, Germany, United States, Slovenia, Estonia, Malta |

TN3 and TN4 represent a lower-intermediate tier and display a more heterogeneous composition. TN3 mainly groups countries from North Africa, the Middle East, and some

cases from the post-Soviet or Eurasian periphery, along with selected countries from Latin America and Africa. This group is characterized by medium income levels but lower generosity and not distinctively high SOCSUP. TN4 includes a mix of South and Southeast Asia, emerging Sub-Saharan Africa, and some lower-income Latin American countries, forming a block of developing economies with relative improvements but still far from high happiness levels.

In the intermediate range, TN5 groups a combination of countries from the Eastern and Southeastern Mediterranean, Eastern Europe, and the post-Soviet space, along with some cases from southern Africa and crisis-affected Latin America. It shows relatively high SOCSUP levels, although with more uneven economic and health trajectories. TN6 is dominated by Latin America, along with several countries from Eastern Europe/Balkans and Central and Southeast Asia. This node reflects middle-income economies with relatively high SOCSUP and higher FREE, supporting intermediate-to-high levels of happiness.

In the upper range, TN7 mainly includes more stable Latin American countries, along with economies from the Balkans, Central Asia, Southeast Asia, and some cases from Eastern Europe. It represents a middle-to-upper-income group where the combination of strong social support, greater freedom, and better health conditions is clearly associated with high levels of happiness. TN8, by contrast, has a more defined profile of Southern Europe and Central and Eastern Europe integrated into the EU, along with some developed East Asian economies. Countries such as Spain, Italy, Portugal, France, and Japan appear here, suggesting a block of advanced economies with high life expectancies. Indeed, life expectancy in Spain, Italy, and Japan is consistently among the highest in the world.

Above this group, TN9 includes highly prosperous and open economies, many of which are small or highly internationalized, including wealthy Western European countries and Gulf states, as well as global hubs in Asia. Finally, TN10 concentrates the group with the highest levels of happiness and presents a very clear identity: Nordic countries predominate, together with advanced economies from Western and Central Europe, as well as developed Anglosphere countries such as Australia, New Zealand, Canada, and the United States. This node can be interpreted as the block of advanced democracies with the highest combined levels of SOCSUP, GDP, HEALTH, and FREE.

Overall, the nodes suggest a fairly recognizable geography of well-being: at the lower end, Sub-Saharan Africa and fragile countries in Asia and the Middle East predominate; in the intermediate range, Latin America, the Balkans, Eastern Europe, and emerging Asia appear; and at the upper end, advanced Europe, the Nordic countries, developed Anglosphere economies, and some high-income hubs in the Gulf and Asia are concentrated.

### *4.2. Results of the Necessary Condition Analysis (Research Objective 2)*

Table 8 reports the effect size (d) of the necessary condition status for all the variables included in the analysis. The SOCSUP has a large effect, since $d > 0.3$ (Dul et al., 2023), under both methods tested ($d = 0.330$ in CE-FDH and 0.319 in CR-FDH), and this effect is statistically significant ($p < 0.001$).

GDP, HEALTH, and FREE exhibit medium effect sizes ($0.1 < d < 0.3$), which are also significant ($p < 0.001$), regardless of the ceiling envelopment method used. Notably, FREE, although it ranks as the fourth most important variable in the decision tree, has the second highest effect in terms of necessary condition status. In fact, its effect size is considerably larger than that of HEALTH and GDP.

In contrast, both GENER and CORRUP consistently have $d < 0.1$, although their effects remain statistically significant. In the most conservative case, when CR-FDH is used, GENER is $d = 0.038$ ($p = 0.040$), whereas CORRUP is $d = 0.050$ ($p = 0.007$).

**Table 8.** Effect sizes of the necessary condition status of the explanatory variables.

| Variable | Effect Size (d) (CE-FDH) | Effect Size (d) (CR-FDH) |
|---|---|---|
| GDP | 0.206 (<0.001) | 0.197 (<0.001) |
| SOCSUP | 0.330 (<0.001) | 0.319 (<0.001) |
| HEALTH | 0.167 (<0.001) | 0.155 (<0.001) |
| FREE | 0.255 (<0.001) | 0.229 (<0.001) |
| GENER | 0.040 (0.001) | 0.038 (0.040) |
| CORRUP | 0.066 (<0.001) | 0.050 (0.007) |

Note: In parentheses are *p* values. CE-FDH = Ceiling Envelopment-Free Disposal Hull, and CR-FDH = Ceiling Regression-Free Disposal Hull.

Table 9 presents the bottleneck table for the CE-FDH estimation, while Figures 5–7 show the NCA plots of HAPPY with the explanatory variables, allowing visualization of the free disposal hulls. Taken together, these results make it possible to interpret the NCA effect sizes beyond merely mechanical reading. The measure *d* should be understood in conjunction with the bottleneck table and the broader tree-based and correlational results. In this sense, the SOCSUP stands out because it already appears as a bottleneck at relatively low and intermediate happiness thresholds. Compared with GDP or HEALTH, FREE is also noteworthy since its bottleneck role appears earlier and more consistently.

**Table 9.** Bottleneck table for the ceiling envelopment-free disposal hull.

| Accumulated | HAPPY | GDP | SOCSUP | HEALTH | FREE | GENER | CORRUP |
|---|---|---|---|---|---|---|---|
| 0 | 1.28 | NN | NN | NN | NN | NN | NN |
| 10% | 1.93 | NN | 0.28 | NN | 0.32 | NN | NN |
| 20% | 2.58 | NN | 0.28 | NN | 0.32 | NN | NN |
| 30% | 3.23 | NN | 0.28 | NN | 0.32 | NN | NN |
| 40% | 3.88 | NN | 0.28 | NN | 0.32 | NN | NN |
| 50% | 4.53 | NN | 0.43 | NN | 0.32 | NN | NN |
| 60% | 5.18 | NN | 0.51 | 47.88 | 0.41 | NN | NN |
| 70% | 5.83 | 9.39 | 0.66 | 52.54 | 0.41 | −0.30 | NN |
| 80% | 6.48 | 9.85 | 0.72 | 55.70 | 0.59 | −0.28 | NN |
| 90% | 7.13 | 10.34 | 0.79 | 60.32 | 0.65 | −0.18 | 0.686 |
| 100% | 7.78 | 11.16 | 0.89 | 60.32 | 0.80 | −0.05 | 0.106 |

Note: Quantities are actual variables, and NN stands for not necessary.




**Figure 5.** Ceiling-Envelopment free-disposal hull of HAPPY with GDP and SOCSUP.

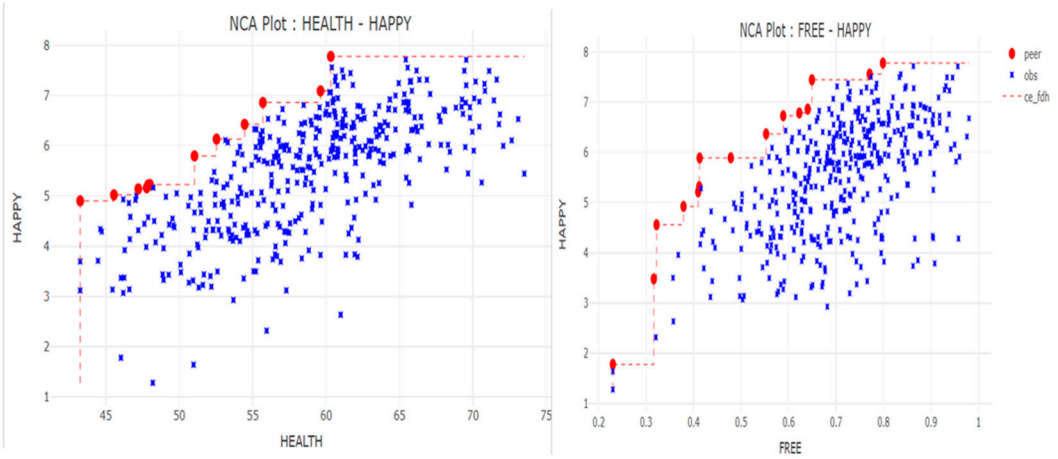



**Figure 6.** Ceiling-Envelopment-free-disposal hull of HAPPY with HEALTH and FREE.

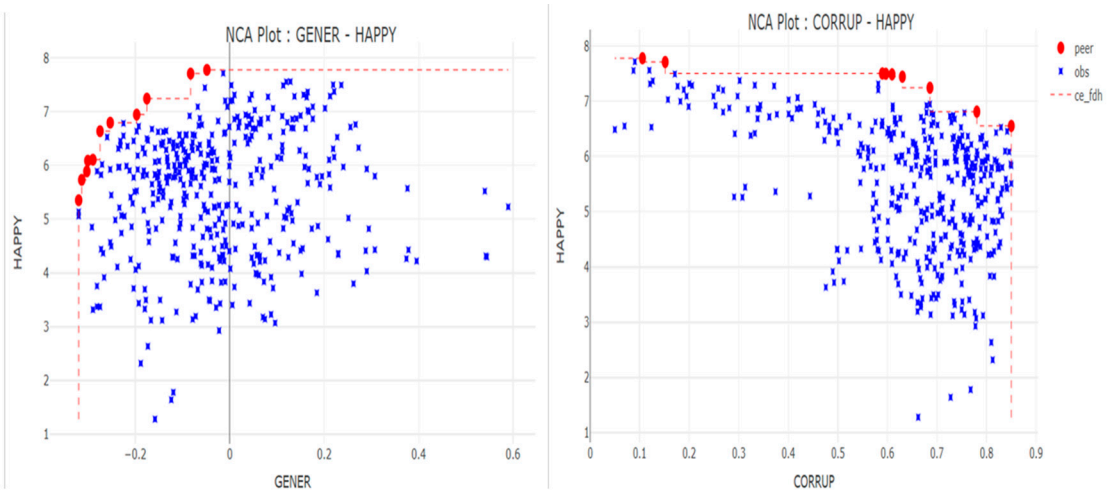



**Figure 7.** Ceiling-Envelopment free-disposal hull of HAPPY with GENER and CORRUP.

At relatively low HAPPY levels, between the 10th percentile (1.93) and the 30th percentile (3.23), SOCSUP and FREE already emerge as bottlenecks. More specifically, from the 10th to the 40th percentile of HAPPY, the estimated thresholds remain approximately constant at 0.28 for SOCSUP and 0.32 for FREE, out of a maximum possible value of 1 for both variables. At the 50th percentile, the required value of SOCSUP approaches 0.5, while FREE remains at 0.32. As happiness levels increase, so do the required thresholds for both variables: to reach 80% of HAPPY, SOCSUP must be 0.72 and FREE 0.59; finally, to approach the highest levels of happiness, SOCSUP must reach 0.89 and FREE 0.8.

HEALTH and GDP, by contrast, do not begin to act as bottlenecks until HAPPY levels increase: approximately 60% (5.18) for HEALTH and 70% (5.73) for GDP. These results suggest that relatively high levels of happiness are observed even with low GDP when accompanied by high levels of SOCSUP and FREE. In fact, achieving high levels of happiness does not appear to require exceptionally high healthy life expectancy, as approximately 60 years would suffice, even though the maximum value in Table 2 for the sample exceeds 70 years.

With respect to GENER and CORRUP, both variables become bottlenecks at higher quantiles: approximately 70% in the case of generosity and 90% in the case of corruption. Notably, to reach the 90% or 100% level of happiness, positive values of GENER do not appear necessary within NCA. In other words, maximum levels of happiness can be observed even when recorded generosity falls below what would be expected on the basis of a country's GDP.

Finally, regarding the role of CORRUP, it should be emphasized that although it does not constitute a constraint for achieving high levels of HAPPY—even at 80% (equivalent to a CL value of 6.48)—it becomes relevant when aiming for very high levels of happiness. Thus, to reach 90% of HAPPY (7.13), it is necessary that perceived corruption is not excessive (CORRUP = 0.686), whereas happiness levels close to the maximum are associated with very low levels of perceived corruption (0.106).

Table 10 presents the Spearman correlation across the different analytical techniques used to assess the importance of variables in relation to HAPPY (correlation analysis, decision tree importance, and effect size of variables as bottlenecks). Although this correlation is high, it is far from perfect, particularly when comparing correlation analysis and the decision tree with NCA. Thus, a variable of secondary relevance according to the decision tree, such as FREE, nonetheless emerges as the second most important bottleneck for achieving a high level of societal well-being, ranking above both income and life expectancy.

**Table 10.** Spearman correlation of factor rankings across the three methodologies used.

| | **Pearson's Correlation** | **Decision Tree** | **NCA** |
|---|---|---|---|
| Pearson's correlation | 1 | | |
| Decision Tree | 0.943 | 1 | |
| NCA | 0.829 | 0.771 | 1 |

## 5. Discussion

### *5.1. General Considerations*

The results obtained in this study contribute to advancing the understanding of happiness (HAPPY) within the framework of the baseline World Happiness Report (WHR) model (Helliwell et al., 2025) from an analytical perspective that differs from the dominant approach in the literature, which is traditionally based on linear regression models. Accordingly, the analytical framework adopted considers happiness in relation to gross domestic product per capita (GDP), social support (SOCSUP), healthy life expectancy (HEALTH), freedom to make life choices (FREE), generosity (GENER), and perceptions of corruption (CORRUP). In line with the stated objectives, the combined use of decision tree regression (DTR) and necessary condition analysis (NCA) suggests that observed happiness levels are not well captured by simple linear relationships but are instead associated with the interaction of factors and structural constraints.

In response to the first research objective (RO1), the DTR results highlight SOCSUP and GDP as the main structural axes in the tree-based analysis of happiness from a logical perspective of sufficiency. These findings are consistent with the revised WHR literature, whose mainstream work identifies both variables as the main correlates of subjective well-being (Chaudhary et al., 2020; Helliwell et al., 2025; Tofallis, 2020; Zhang, 2023). However, the approach adopted here goes beyond this general observation by showing that these variables are not only strongly linked to happiness on average but also structure the segmentation of countries into differentiated well-being regimes. In particular, the fact that the first split of the tree occurs on social support suggests that this dimension operates as an initial partitioning condition across societies, shaping different "possibility spaces" within which the remaining variables operate. Moreover, all the identified groups (up to 10) emerge from the interaction between SOCSUP and GDP as primary splits. Depending on the terminal node, variables such as FREE, GENER, and CORRUP also play a role. Thus, this is in line with previous WHR-based research identifying GDP and SOCSUP as central

correlates of subjective well-being, but our results add a configurational nuance by showing that these variables also structure differentiated well-being regimes across countries.

The prominence of GDP and the SOCSUP can also be justified theoretically. Following the distinction between life chances and life results, happiness as life satisfaction is better understood as an outcome of life, whereas the surrounding conditions that make such outcomes more or less likely should be interpreted as broad enabling bases rather than as direct or isolated causes (Veenhoven, 2025). From this perspective, GDP per capita may be understood as a foundational material condition of well-being, since it reflects the resource base that expands security, reduces deprivation, and enlarges the feasible set of life opportunities. However, the happiness literature has long shown that the positive role of income is limited and tends to weaken over time because gains in income are partly offset by social comparison and adaptation processes. GDP is therefore better interpreted as an enabling condition that helps set the material floor of well-being than as a sufficient engine of sustained increases in happiness (Easterlin & O'Connor, 2020). In contrast, social support captures the relational infrastructure of well-being—trust, care, companionship, and effective support in times of need—and evidence from the WHR highlights social connection as central to subjective well-being across multiple domains of social life (Helliwell et al., 2025). In this sense, GDP and social support can be seen as two broad supports of a good life within the WHR framework: the first is primarily material, and the second is relational.

The DTR results highlight the importance of interactions among variables in explaining happiness and support previous findings in the literature. Unlike linear models, where effects are interpreted independently unless interactions are explicitly specified, as in Kamińska et al. (2025), the tree shows that the association of variables such as freedom, generosity, or corruption depends on the levels reached in structural variables such as social support or income. This finding reinforces the idea, widely discussed in the recent literature, that happiness responds to multiple configurations (Pereira et al., 2023), with configurational approaches showing that multiple paths can lead to a given level of happiness.

A significant finding is the role of variables with low apparent correlation, such as GENER, which nonetheless emerges as relevant in the decision tree. This result has a twofold interpretation. On the one hand, it highlights the limitations of traditional correlation analysis, since variables with weak linear relationships may still play a meaningful role in specific contexts. On the other hand, although the fitted decision tree has shown robustness, at least in its upper branches, this pattern should be interpreted with caution because the appearance of GENER in lower-level splits may also reflect local interaction structures rather than broad substantive importance across the full sample. The weak linear but context-dependent role of GENER is compatible with contributions that question its broad explanatory relevance in standard happiness models while still leaving room for more localized or conditional associations in specific empirical settings (e.g., Zong, 2024).

The classification of countries into terminal nodes reveals a geography of well-being consistent with existing empirical evidence (Çelik et al., 2025b; Detrinidad & López-Ruiz, 2024; Helliwell et al., 2025), enriched by an interpretation of the combinations of factors that act as underlying positively linked correlates. The identified groups make it possible to distinguish countries not only by their levels of happiness but also by the way this outcome is related to different combinations of explanatory factors.

The development of RO2 through NCA introduces a novel perspective by analyzing which WHR variables emerge as necessary conditions and, therefore, as potential bottlenecks from an NCA perspective in the development of societal well-being. The results show that the SOCSUP is not only the most prominent variable in the tree-based analysis

but also the strongest necessary condition for attaining high levels of happiness. Similarly, FREE emerges as a relevant necessary condition, even above GDP or HEALTH, including at relatively low levels of HAPPY. These results suggest that certain levels of individual autonomy appear to constitute a prerequisite within NCA for attaining even modest levels of happiness. The strong relevance of FREE is consistent with some findings in the literature that, although not dominant, are far from anecdotal (Çelik et al., 2025b; Machani & Shtudiner, 2026; Pološki Vokić & Klindžić, 2024). Thus, these findings on FREE as a bottleneck extend this literature by suggesting that it is not only positively associated with happiness in some settings but also operates as a threshold-like condition.

The patterns described above should nevertheless be interpreted with caution. In a cross-country setting, variables such as GDP, SOCSUP, HEALTH, and FREE are partly intertwined and may reflect broader shared development or institutional environments. Therefore, the NCA results do not imply that each variable operates as an independent causal condition in a strong substantive sense. Rather, they indicate that when each variable is considered separately, higher levels of happiness are not observed below certain empirical thresholds.

In the case of happiness, the SOCSUP proves more relevant than GDP does, and when the level of necessity is analyzed, FREE emerges as a more demanding bottleneck than GDP does; this can be understood by considering that happiness reflects the subjective satisfaction linked to hedonic utility, which, although related to it, constitutes a distinct dimension of utility from instrumental utility. While the former is linked to subjective well-being, the latter is more closely associated with objective and material benefits (Di Foggia et al., 2025).

GENER and CORRUP should be interpreted with particular caution. Although both variables are statistically significant in the NCA estimations, their effect sizes remain below the conventional threshold for practical relevance ($d < 0.1$) (Dul et al., 2023). This suggests that, in overall terms, they do not constitute strong bottlenecks for happiness attainment. Their relevance appears only in a narrow sense, namely, when very high happiness levels are considered. Accordingly, these variables are better interpreted as secondary or context-specific constraints in the upper tail of the happiness distribution rather than as substantively important necessary conditions across the full range of observed happiness levels.

The interpretation of these results should also take into account the limitations of the underlying data. First, happiness is operationalized through the CL, which captures a self-reported evaluative assessment of life rather than the full multidimensional content of well-being (De Paola & Pirttilä-Backman, 2022). Second, some explanatory variables, particularly generosity, are constructed proxies from residuals (Helliwell et al., 2025) and may therefore contain substantial measurement noise. Third, the period analyzed (2022–2024) is relatively short and may partly reflect postpandemic dynamics rather than fully stable long-term relationships across countries.

### *5.2. Theoretical and Practical Implications*

The results obtained through the decision tree provide a very useful complement to the conventional regression approach. While traditional regression facilitates the identification of statistically significant relationships, the tree makes it possible to empirically examine how the most relevant variables interact to explain specific levels of happiness. In addition, the level of fit achieved is very similar to that reported in regression-based approaches, which typically show an $R^2$ close to 80% (Çelik et al., 2025b; Helliwell et al., 2025), reinforcing the validity of DTR. In this respect, our approach is close to recent studies that apply machine-learning tools to happiness data (Çelik et al., 2025a; Chaudhary

et al., 2020; Khder et al., 2022; Timmapuram et al., 2023), but it differs from predominantly predictive uses of these methods by emphasizing interpretability, threshold structure, and country segmentation.

Second, the incorporation of NCA represents a noteworthy contribution to the theoretical debate. While most studies focus on sufficiency relationships (i.e., which factors increase happiness), NCA introduces a different logic aimed at identifying minimum requirements in an NCA setting. This shift in perspective is crucial, as it implies that some variables appear not to be easily compensated for by others. For example, the NCA results suggest that higher income does not empirically compensate for very low levels of social support or freedom, which calls into question simplified approaches based on economic growth as the primary path to well-being, as suggested by the Easterlin paradox (Easterlin, 1974).

To illustrate how the bottleneck information can be used in practice, let us consider a country whose HAPPY = 5.165, GDP = 9.901, SOCSUP = 0.835, HEALTH = 47.78, FREE = 0.728, GENER = −0.182, and CORRUP = 0.790. Suppose that a policymaker wishes to identify the main variables on which to focus to improve citizens' happiness. In light of the bottleneck results reported in Table 9, HEALTH appears to be the most immediate policy-relevant bottleneck in this illustrative case. Indeed, to reach the 60% happiness threshold (HAPPY = 5.18), this country already satisfies the requirements for SOCSUP and FREE, but it does not meet the minimum required level of health (HEALTH = 47.88), which therefore emerges as its immediate bottleneck. The same pattern remains in place for further progress, since at the 70% (HAPPY = 5.83) and 80% (HAPPY = 6.48) thresholds, HEALTH is still the only bottleneck that is not satisfied. In contrast, approaching the 90% happiness threshold (HAPPY = 7.13) would require not only improvements in health but also the simultaneous fulfillment of several additional thresholds: GDP would need to rise to 10.34 (from 9.90), HEALTH to 60.32 (from 47.78), GENER to −0.18 (from −0.182, thus with only a marginal gap), and CORRUP would need to decline to a maximum of 0.686 (from 0.790). Taken together, these results suggest that while modest gains in happiness may primarily depend on improvements in health, progress toward much higher happiness levels would be associated with a broader set of unmet thresholds, particularly income, health, and perceived corruption.

The results also reinforce the idea of happiness as a multidimensional and nonlinear phenomenon, aligning with recent approaches that emphasize equifinality and causal asymmetry (Pereira et al., 2023). In this sense, the study contributes to consolidating a theoretical framework in which happiness is not well captured by a single universal function but rather by multiple possible configurations of factors.

From a practical perspective, the implications are equally relevant. First, the results suggest that public policies aimed at improving well-being should not focus exclusively on economic growth but should prioritize strengthening social capital. The SOCSUP emerges as the most important factor in NCA, suggesting that policies promoting social cohesion, interpersonal trust, and support networks may be notably connected to higher happiness levels. Nor should the need for more comprehensive family policies be overlooked—policies aimed at reducing the costs of forming a family and creating a social environment more conducive to marriage and childrearing (Choi et al., 2024).

The identification of minimum thresholds through the NCA provides a useful tool for policy design. Rather than seeking marginal improvements across all the factors, policymakers can focus on ensuring minimum levels in key variables. For example, guaranteeing certain levels of FREE or SOCSUP may be more relevant than marginal increases in GDP in contexts where these conditions are not met.

Thus, from a policy perspective, the finding that FREE emerges as a relevant bottleneck should not be interpreted as evidence of a simple causal effect whereby increasing this vari-

able would mechanically increase happiness. Rather, from an NCA perspective (Dul, 2016), very low levels of perceived freedom to make life choices may constrain the attainment of higher happiness levels, even when other conditions are relatively favorable. In practical terms, this points to the importance of not relying exclusively on income-oriented policies. Interventions that strengthen people's effective room for choice, such as through better institutional quality, greater personal autonomy, reduced administrative constraints, and more secure life opportunities, may be especially relevant in contexts where this dimension remains weak. Moreover, these implications should be treated with caution, since FREE is a broad country-level indicator and may capture several underlying institutional and social mechanisms rather than a single directly actionable policy instrument.

Finally, the segmentation of countries into homogeneous groups suggests that policies should be tailored to specific contexts. The results do not support a single strategy applicable to all countries; instead, interventions must be designed taking into account the structural configuration of each case. This is particularly relevant for developing countries, where improvements in well-being are more closely tied to social and institutional factors than to increases in income.

## 6. Conclusions

### *6.1. Principal Findings*

In this study, the factors related to happiness are analyzed on the basis of the WHR model, and a methodological approach based on decision trees and necessary condition analysis is introduced. The results allow several main conclusions to be drawn. In answer to the central research question, the evidence suggests that the core WHR variables are not only associated with happiness through interaction patterns but also that some of them—most notably social support and, to a lesser extent, freedom in the NCA setting—also emerge as threshold-like conditions for attaining higher observed happiness levels.

The central role of social support and GDP per capita as structuring variables associated with happiness is supported. However, their importance is not limited to average effects; rather, both emerge as the main partitioning variables in the tree.

The findings show that observed happiness levels are connected to complex configurations of factors characterized by nonlinearities and interactions. Variables such as generosity, which exhibit a weak connection with happiness, may still play a relevant role in specific contexts, highlighting the need for more flexible analytical approaches beyond traditional regression models.

NCA shows that SOCSUP and FREE emerge as demanding bottlenecks for achieving high levels in HAPPY. This result has important interpretive implications, as it indicates that it is difficult for certain factors to substitute for others in NCA.

Overall, the results suggest that happiness should be understood as a multidimensional, nonlinear phenomenon shaped by minimum thresholds, which suggests the usefulness of complementary analytical approaches and more context-sensitive policy interpretations.

### *6.2. Limitations and Further Research*

This study presents several limitations that should be considered when interpreting the results and that open new avenues for future research. First, from a methodological perspective, the decision tree model, although highly interpretable, has limitations in terms of stability and sensitivity to the sample. Small variations in the data may generate different tree structures, which can affect the robustness of the results. Similarly, the restriction imposed to ensure interpretability (limiting the number of nodes) involves a trade-off with the predictive capacity of the model. In any case, the robustness checks conducted in this study suggest that, in our application, the upper layers of the tree are highly stable and

that the ranking of the main associations remains largely unchanged across alternative specifications and resampling procedures.

While offering a novel perspective, the NCA depends on data quality and the correct specification of variables. An additional methodological caveat is that several WHR variables are correlated. Although NCA is designed to assess each condition separately, these correlations imply that the identified bottlenecks should not be interpreted as empirically independent conditions in a strong causal sense. Future research could examine the stability of these necessity patterns using longer time spans, alternative operationalizations, and designs combining NCA with multivariate or longitudinal approaches.

Another important limitation relates to the WHR model itself. This framework reduces happiness to an overly simplified measure, largely based on cognitive evaluations of life, without adequately capturing its emotional and cultural dimensions. As a result, it may fall short of accurately reflecting the lived social reality of individuals (De Paola & Pirttilä-Backman, 2022). That said, the analytical strategy presented in this study can be readily extended to incorporate new explanatory variables across all stages of the proposed evaluation.

The measurement of happiness through the CL also presents conceptual limitations. It is based on a single-item measure, which implies a simplification of a complex and multidimensional phenomenon. This measure may be influenced by contextual, cultural, or momentary factors and does not fully capture dimensions such as emotional well-being, sense of purpose, or satisfaction across different domains of life. Although widely used in the literature, responses may still be influenced by cultural differences in self-evaluations, response styles, and the subjective interpretation of what constitutes the "best possible life" or the "worst possible life" (Nilsson et al., 2024).

In addition, generosity is a constructed variable, operationalized as a residual rather than as a direct measure, which may introduce substantial measurement noise and complicate its interpretation. This is relevant when assessing its role in nonlinear splits or as a potential necessary condition. Future research could address this issue by testing alternative operationalizations of generosity and examining whether its role remains stable across different measurement strategies.

An additional limitation is that the period analyzed (2022–2024) is relatively short and may partly reflect postpandemic dynamics. Therefore, some of the observed cross-country patterns may capture short- to medium-term adjustments rather than fully stable long-term relationships. Future research could address this issue by comparing prepandemic and postpandemic periods to assess the extent to which the identified patterns remain stable over time.

Future research could extend the analysis by incorporating multidimensional measures of well-being and compare the present findings with those of other ensemble and longitudinal modeling strategies. More broadly, integrating institutional, cultural, or inequality-related variables could enrich the understanding of the phenomenon. Finally, the development of hybrid approaches that combine the logic of necessity, sufficiency, and causality represents a promising research agenda for advancing the study of happiness.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/economies14050172/s1, The annotated R scripts used in the empirical analysis are also publicly available in the GitHub repository associated with this study (https://github.com/Falarita/Explaining-Global-Happiness-Evidence-from-Decision-Trees-and-Necessary-Condition-Analysis, accessed on 24 April 2026).

**Author Contributions:** Conceptualization, T.T.-C. and J.d.A.-S.; methodology, T.T.-C. and J.d.A.-S.; software, T.T.-C. and J.d.A.-S.; validation, T.T.-C. and J.d.A.-S.; formal analysis T.T.-C. and J.d.A.-S.; investigation, T.T.-C. and J.d.A.-S.; resources, T.T.-C. and J.d.A.-S.; data curation, T.T.-C. and J.d.A.-S.;

writing—original draft preparation, T.T.-C.; writing—review and editing, J.d.A.-S.; visualization, T.T.-C. and J.d.A.-S.; supervision, T.T.-C. and J.d.A.-S.; project administration, T.T.-C. and J.d.A.-S.; funding acquisition, T.T.-C. and J.d.A.-S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Telefonica and the Telefonica Chair on Smart Cities of the Universitat Rovira i Virgili and Universitat de Barcelona (project number: 42. DB.00.18.00).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available through https://www.worldhappiness.report/data-sharing/ (accessed on 25 February 2026).

**Acknowledgments:** ChatGPT 5.5. has been used to improve the linguistic quality of the paper.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

| | |
|---|---|
| WHR | World Happiness Report |
| CL | Cantril Ladder |
| HAPPY | Happiness measured through Cantril Ladder |
| GDP | Gross Domestic Product per capita |
| SOCSUP | Social Support |
| HEALTH | Healthy Life Expectancy |
| FREE | Freedom to make life choices |
| GENER | Generosity |
| CORRUP | Perceived Corruption |
| DTR | Decision Tree Regression |
| NCA | Necessary Condition Analysis |
| CE-FDH | Ceiling Envelopment-Free Disposal Hull |
| CR-FDH | Ceiling Regression-Free Disposal Hull |
| RMSE | Root Mean Squared Error |
| MAE | Mean Absolute Error |
| $R^2$ | Coefficient of Determination |

## References

Agrawal, S., Sharma, N., Bruni, M. E., & Iazzolino, G. (2023). Happiness economics: Discovering future research trends through a systematic literature review. *Journal of Cleaner Production*, *416*, 137860. [CrossRef]

Akanbi, K., Jones, Y., Oluwadare, S., & Nti, I. K. (2024, April 13–14). *Predicting happiness index using machine learning*. 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI), Mt Pleasant, MI, USA. [CrossRef]

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. [CrossRef]

Chaudhary, M., Dixit, S., & Sahni, N. (2020). Network learning approaches to study world happiness. *arXiv*. [CrossRef]

Chen, X., Li, S., Sowah, J. S., & Zou, S. (2025). Soft institutions, hard cash: Societal happiness and corporate cash holdings. *International Review of Economics & Finance*, *103*, 104506. [CrossRef]

Choi, K., Kim, G., Yoo, D., & Lee, J. (2024). Does economic stability influence family development? Insights from women in Korea with the lowest childbirth rates worldwide. *Economies*, *12*(3), 74. [CrossRef]

Clarke, S. (2022). Freedom and happiness: Does freedom make people happy? *Journal of Political Science: Bulletin of Yerevan University*, *1*(1), 140–155. [CrossRef]

Collado, Z. C. (2024). Does the intensity of conflict-induced internal displacement influence national happiness scores? *The Social Science Journal*, 1–9. [CrossRef]

Çelik, S., Doğan, İ., Uçal, M., & Akbulut Küçük, S. (2025a). Accuracy comparison of machine learning algorithms on World Happiness Index data. *Mathematics*, *13*(3), 1176. [CrossRef]

Çelik, S., Gökdemir, T., & Çondur, F. (2025b). Statistical analysis of the world happiness index by regions. *Quality & Quantity*, *59*, 4003–4018. [CrossRef]

De Paola, J., & Pirttilä-Backman, A.-M. (2022). Are we bad winners? Public understandings of the United Nations' world happiness report among finnish digital media and their readers. *Public Understanding of Science*, *32*(1), 20–39. [CrossRef]

Detrinidad, E., & López-Ruiz, V.-R. (2024). The interplay of happiness and sustainability: A multidimensional scaling and K-means cluster approach. *Sustainability*, *16*, 10068. [CrossRef]

Di Foggia, G., Arrigo, U., & Beccarello, M. (2025). Evolution and theoretical implications of the utility concept. *Economies*, *13*(10), 283. [CrossRef]

Dul, J. (2016). Necessary condition analysis (NCA): Logic and methodology of "necessary but not sufficient" causality. *Organizational Research Methods*, *19*(1), 10–52. [CrossRef]

Dul, J. (2024a). A different causal perspective with necessary condition analysis. *Journal of Business Research*, *177*, 114618. [CrossRef]

Dul, J. (2024b). How to sample in necessary condition analysis (NCA). *European Journal of International Management*, *23*(1), 1–12. [CrossRef]

Dul, J., Hauff, S., & Bouncken, R. B. (2023). Necessary condition analysis (NCA): Review of research topics and guidelines for good practice. *Review of Managerial Science*, *17*(2), 683–714. [CrossRef]

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David, & M. W. Reder (Eds.), *Nations and households in economic growth: Essays in honour of moses Abramovitz* (pp. 89–125). Academic Press. [CrossRef]

Easterlin, R. A., & O'Connor, K. J. (2020). The Easterlin paradox. In K. F. Zimmermann (Ed.), *Handbook of labor, human resources and population economics* (pp. 1–25). Springer International Publishing. [CrossRef]

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research*, *26*(3), 499–510. [CrossRef]

Grinde, B. (2022). Role of happiness when evaluating society. *Encyclopedia*, *2*, 230–236. [CrossRef]

Hauff, S., Guerci, M., Dul, J., & van Rhee, H. (2021). Exploring necessary conditions in HRM research: Fundamental issues and methodological implications. *Human Resource Management Journal*, *31*(1), 18–36. [CrossRef]

Helliwell, J. F., Layard, R., Sachs, J. D., De Neve, J.-E., Aknin, L. B., & Wang, S. (2025). *World happiness report 2025*. Wellbeing Research Centre, University of Oxford.

Hofstede, G. (1980). Culture and organizations. *International Studies of Management & Organization*, *10*(4), 15–41. [CrossRef]

Ionescu-Feleagă, L., Ionescu, B.-Ş., & Stoica, O. C. (2022). The impact of digitalization on happiness: A European perspective. *Mathematics*, *10*(15), 2766. [CrossRef]

Kamińska, J. A., Dionísio, A., Infante, P., & Carrilho, R. (2025). More is still not enough—What is necessary and sufficient for happiness? *Sustainability*, *17*, 6121. [CrossRef]

Kariuki, C. W., & Karanu, J. W. (2024). The relationship between happiness and foreign direct investment in African countries. *Economies*, *12*(12), 343. [CrossRef]

Kasman, S., & Kasman, A. (2023). The impact of obesity and income on happiness: Evidence from EU countries. *Panoeconomicus*, *70*(2), 303–319. [CrossRef]

Khder, M. A., Sayfi, M. A., & Fujo, S. W. (2022). Analysis of world happiness report dataset using machine learning approaches. *International Journal of Advance Soft Computing and Applications*, *14*(1), 15–34. [CrossRef]

Lago, I., de Gispert, C., Bosch, N., & Vilalta, M. (2025). Decentralizing happiness. *Publius: The Journal of Federalism*, pjaf012. [CrossRef]

Leavitt, G. (2025). Life, liberty and happiness: In pursuit of the ninth amendment. *American Journal of Society and Law*, *4*(1), 15–25. [CrossRef]

Lee, A. S. H., Kudbuddin, A. B., & Teh, P. L. (2019). Data visualisation: World happiness at a glance! In *Advances in visual informatics* (Vol. 11870). Springer. [CrossRef]

Li, T., Zhong, J., & Xu, M. (2020). Does the credit cycle have an impact on happiness? *International Journal of Environmental Research and Public Health*, *17*(1), 183. [CrossRef] [PubMed]

Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(1), 14–23. [CrossRef]

Machani, M., & Shtudiner, Z. (2026). The effect of corruption priming on happiness and negative emotions. *Managerial and Decision Economics*, *47*, 457–467. [CrossRef]

Mahalik, J. R., & Harris, M. P. (2026). Precarious manhood, precarious nations: The contribution of cultural beliefs comprising masculinity to national happiness. *Social Science & Medicine*, *388*, 118752. [CrossRef]

Morse, S. (2026). Happiness in the sustainable development goals: Adding to the quagmire of quantification? *Sustainability*, *18*(7), 3233. [CrossRef]

Nilsson, A. H., Eichstaedt, J. C., Lomas, T., Schwartz, A., & Kjell, O. (2024). The cantril ladder elicits thoughts about power and wealth. *Scientific Reports*, *14*(1), 2642. [CrossRef] [PubMed]

Organisation for Economic Co-operation & Development. (2015). *How's life? Measuring well-being*. OECD Publishing. [CrossRef]

Pawliczek, A., Kurowska-Pysz, J., & Smilnak, R. (2022). Relation between globe latitude and the quality of life: Insights for public policy management. *Sustainability*, *14*(3), 1461. [CrossRef]

Pereira, M. C., Coelho, F., & Silva, G. M. (2023). Is there a happy culture? Multiple paths to national subjective well-being. *Kyklos*, *76*(4), 613–641. [CrossRef]

Pološki Vokić, N., & Klindžić, M. (2024). The association between national human resource management practices and measures and correlates of national happiness. *Društvena Istraživanja*, *33*, 291–313. [CrossRef]

Powdthavee, N., Burkhauser, R. V., & De Neve, J.-E. (2017). Top incomes and human well-being: Evidence from the gallup world poll. *Journal of Economic Psychology*, *62*, 246–257. [CrossRef]

Rad, D., Cuc, L. D., Croitoru, G., Gomoi, B. C., Mazuru, L., Bilți, R. S., Rusu, S., Sinaci, M., & Barbu, F. S. (2025). Modeling investment decisions through decision tree regression—A behavioral finance theory approach. *Electronics*, *14*(8), 1505. [CrossRef]

Rajkumar, R. P. (2023). Cultural values and changes in happiness in 78 countries during the COVID-19 pandemic: An analysis of data from the world happiness reports. *Frontiers in Psychology*, *14*, 1090340. [CrossRef] [PubMed]

Ram, R. (2021). International convergence in population happiness: Evidence from recent data. *Applied Economics*, *53*(34), 3984–3991. [CrossRef]

Stiglitz, J. E., Sen, A., & Fitoussi, J.-P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Available online: https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf (accessed on 23 December 2025).

Stupak, T. (2024). *How much data is required to train ML models in 2024?* Available online: https://www.akkio.com/post/how-much-data-is-required-to-train-ml (accessed on 25 February 2026).

Timmapuram, M., Ramdas, R., Reddy Vutkur, S., Reddy Mali, Y., & Vasanth, K. (2023, December 14–15). *Understanding the regional differences in world happiness index using machine learning*. 2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES), Chennai, India. [CrossRef]

Tofallis, C. (2020). Which formula for national happiness? *Socio-Economic Planning Sciences*, *70*, 100688. [CrossRef]

Trinh, T. A., & Nhieu, N. L. (2026). A r,s,t-spherical fuzzy decision-making model of university happiness: Case study of University of Economics Ho Chi Minh City. *Humanities and Social Sciences Communications*, *13*, 608. [CrossRef]

Turner, B. S. (2018). (I can't get no) satisfaction: Happiness and successful societies. *Journal of Sociology*, *54*(3), 279–293. [CrossRef]

Veenhoven, R. (2025). How to take stock of research findings on happiness in regions using the world database of happiness. *Applied Research in Quality of Life*, *21*, 9–33. [CrossRef]

Vittersø, J. (2025). Happiness and wellbeing. In *Humanistic wellbeing: Toward a value-based science of the good life* (pp. 49–77). Springer Nature. [CrossRef]

Yang, B., & Xie, X. (2025). Analyzing and predicting global happiness index via integrated multilayer clustering and machine learning models. *PLoS ONE*, *20*(4), e0322287. [CrossRef]

Zhang, Y. (2023). Analyze and predict the 2022 world happiness report based on the past year's dataset. *Journal of Computer Science*, *19*(4), 483–492. [CrossRef]

Zong, Y. (2024, August 9). *The research on the factors affecting the world happiness index*. 2nd International Conference on Mathematical Physics and Computational Simulation, Glasgow, UK. [CrossRef]

Zweig, J. S. (2015). Are women happier than men? Evidence from the gallup world poll. *Journal of Happiness Studies*, *16*, 515–541. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.